
 M O S E N E F
 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Sat May 13 09:43:25 2000; MasPar time 4.90 Seconds
 Tabular output not generated. 333.332 Million cell updates/sec

Title: >US-09-331-631-1
 Description: (117-185) from US09331631.pep (4 of 5)
 Perfect Score: 539
 Sequence: 1 NR0RDPQOQYEOCKHRCOR.....EEOQREDEKYEERKKEEDN 69

Scoring table: PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a:geneseqs5
 1:geneseqp

Statistics: Mean 25.022; Variance 122.235; scale 0.205

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description	Pred. No.
1	539	100.0	666	1	Macadamia integrifolia	2.16e-36
2	520	96.5	625	1	Macadamia integrifolia	9.12e-35
3	517	95.9	666	1	Macadamia integrifolia	1.65e-34
4	217	40.3	525	1	Theobroma cacao antiimi	2.27e-09
5	217	40.3	566	1	Sequence encoded by 67	2.27e-09
6	201	37.3	590	1	Gossypium hirsutum ant	4.42e-08
7	129	23.9	444	1	W90340	1.90e-02
8	129	23.9	524	1	G. max SBP1 protein.	6.38e-02
9	122	22.6	593	1	Zea mays antimicrobial	2.98e-01
10	113	21.0	33	1	G. max truncated SBP2	2.98e-01
11	113	21.0	408	1	G. max SBP2 protein.	2.98e-01
12	113	21.0	489	1	P. falciparum LSA-R-NR	9.71e-01
13	106	19.7	316	1	P. falciparum LSA gene	9.71e-01
14	106	19.7	493	1	Stenocarpus sinuatus a	1.15e+00
15	105	19.5	28	1	Antimicrobial maize pe	1.36e+00
16	104	19.3	35	1	Dirofilaria immitis pa	1.60e+00
17	103	19.1	303	1	Mouse SRI-related prot	2.24e+00
18	101	18.7	395	1	Mouse SH3P7 protein.	5.11e+00
19	96	17.8	433	1	HIV-1 protein express	7.09e+00
20	94	17.4	265	1	Tyrosine kinase associ	7.09e+00
21	94	17.4	450	1	Circumsporozoite (CS)-	8.34e+00
22	94	17.4	559	1	Human regulatory molec	8.34e+00
23	93	17.3	432	1	W93954	

ID	Score	Query Match	Length DB	ID	Description	Pred. No.
24	93	17.3	740	1	R27530	
25	93	17.3	740	1	R68838	
26	92	17.1	562	1	R70491	
27	91	16.9	541	1	R37148	
28	91	16.9	765	1	R75915	
29	91	16.9	765	1	R92275	
30	91	16.9	783	1	R37151	
31	91	16.9	787	1	R37152	
32	91	16.9	802	1	R37153	
33	90	16.7	482	1	R05766	
34	89	16.5	482	1	R90546	
35	89	16.5	301	1	R70867	
36	89	16.5	1088	1	R19786	
37	89	16.5	1312	1	R22775	
38	89	16.5	1312	1	R71295	
39	89	16.5	1752	1	R07031	
40	88	16.3	436	1	R03662	
41	88	16.3	614	1	R82630	
42	87	16.1	388	1	R04998	
43	87	16.1	462	1	R05766	
44	86	16.0	376	1	R05137	
45	86	16.0	796	1	R36024	

ALIGNMENTS

ID	Score	Query Match	Length DB	ID	Description	Pred. No.
1	539	100.0	666	1	Macadamia integrifolia	2.16e-36
2	520	96.5	625	1	Macadamia integrifolia	9.12e-35
3	517	95.9	666	1	Macadamia integrifolia	1.65e-34
4	217	40.3	525	1	Theobroma cacao antiimi	2.27e-09
5	217	40.3	566	1	Sequence encoded by 67	2.27e-09
6	201	37.3	590	1	Gossypium hirsutum ant	4.42e-08
7	129	23.9	444	1	W90340	1.90e-02
8	129	23.9	524	1	G. max SBP1 protein.	6.38e-02
9	122	22.6	593	1	Zea mays antimicrobial	2.98e-01
10	113	21.0	33	1	G. max truncated SBP2	2.98e-01
11	113	21.0	408	1	G. max SBP2 protein.	2.98e-01
12	113	21.0	489	1	P. falciparum LSA-R-NR	9.71e-01
13	106	19.7	316	1	P. falciparum LSA gene	9.71e-01
14	106	19.7	493	1	Stenocarpus sinuatus a	1.15e+00
15	105	19.5	28	1	Antimicrobial maize pe	1.36e+00
16	104	19.3	35	1	Dirofilaria immitis pa	1.60e+00
17	103	19.1	303	1	Mouse SRI-related prot	2.24e+00
18	101	18.7	395	1	Mouse SH3P7 protein.	5.11e+00
19	96	17.8	433	1	HIV-1 protein express	7.09e+00
20	94	17.4	265	1	Tyrosine kinase associ	7.09e+00
21	94	17.4	450	1	Circumsporozoite (CS)-	8.34e+00
22	94	17.4	559	1	Human regulatory molec	8.34e+00
23	93	17.3	432	1	W93954	

OS Macadamia integrifolia.
FH Key Location/Qualifiers
FT Peptide 1..28
FT /note="signal peptide"
FT 29..666
FT Protein /note="mature protein"
PN WO9827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997; AU0874.
PR 20-DEC-1996; AU-004275.
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI: 98-377279/32.
DR N-PSDB: V42316.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
PS Claim 1; Page 43-45; 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 625 AA;
Query Match 96.5%; Score 520; DB 1; Length 625;
Best Local Similarity 95.7%; Pred. No. 9.12e-35;
Matches 66; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
DB 76 NR0RPOQYECQRCORRETEPRHMQCQRCERREKRRKQKRYEQOQREDEKX 135
QY 117 NR0RPOQYECQRCORRETEPRHMQCQRCERREKRRKQKRYEQOQREDEKX 176
DB 136 ERMKEEDN 144
QY 177 ERMKEEDN 185
RESULT 3
ID W62829 standard; Protein: 666 AA.
AC W62829.
DT 27-OCT-1998 (first entry)
DE Macadamia integrifolia antimicrobial protein.
KW antimicrobial protein; infestation; control.
OS Macadamia integrifolia.
FH Key Location/Qualifiers
FT Peptide 1..28
FT /note="signal peptide"
FT 29..666
FT Protein /note="mature protein"
PN WO9827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997; AU0874.
PR 20-DEC-1996; AU-004275.
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI: 98-377279/32.
DR N-PSDB: V42316.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
PS Claim 1; Page 39-41; 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 666 AA;
Query Match 95.9%; Score 517; DB 1; Length 666;
Best Local Similarity 95.7%; Pred. No. 1.65e-34;
Matches 66; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 177 ERMKEEDN 185
RESULT 4
ID W62831 standard; Protein: 525 AA.
AC W62831.
DT 27-OCT-1998 (first entry)
DE Theobroma cacao antimicrobial protein.
KW antimicrobial protein; infestation; control.
OS Theobroma cacao.
PN WO9827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997; AU0874.
PR 20-DEC-1996; AU-004275.
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI: 98-377279/32.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
PS Claim 1; Page 47-49; 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 525 AA;
Query Match 40.3%; Score 217; DB 1; Length 525;
Best Local Similarity 47.0%; Pred. No. 2.27e-09;
Matches 31; Conservative 15; Mismatches 16; Indels 4; Gaps 3;
DB 35 ERDPRQYECQRCRESATEEREOECQRCERY-KEQOROE--BELORQYQ-CQG 90
QY 119 ORDPQOYECQRCQRCQRETEPRHMQCQRCERREKRRKQKRYEQOQREDEKX 178
DB 91 RCOEQO 96
QY 179 RRMKEED 184
RESULT 5
ID R20181 standard; Protein: 566 AA.
AC R20181.
DT 16-APR-1992 (first entry)
DE Sequence encoded by 67 kD T. cacao protein cDNA.
KW Cocoa; Flavour; vicillin; seed storage protein.
OS Theobroma cacao.
PN WO911801-A.
PD 26-DEC-1991.
PF 07-JUN-1991; G00914.
PR 11-JUN-1990; GB-013016.
PA (MRSC) MARS UK LTD.
PI Spencer ME, Hodge R, Deakin EA, Ashton S;
DR WPI: 92-024418/03.
DR N-PSDB: Q20377.
PT Recombinant cocoa proteins - are responsible for flavour in cocoa
PT beans and produced in large quantities using yeast and bacterial
PT expression vectors
PS Claim 4; Fig 2; 59pp; English.
CC The inventors claim a 67 kD and 31 kD T. cacao protein, and
CC fragments, and encoding DNAs. The 47 kD and 31 kD proteins are
CC derived from the 67 kD precursor. T. cacao protein cDNA was
CC detected in a cDNA library prepared from immature cocoa beans RNA
CC using a probe based on the AA sequence of a CNBR peptide common to
CC the 47 kD and 31 kD polypeptides. Homology searches revealed close
CC homologues between the 67 kD polypeptide and the vicillins, which are
CC seed storage proteins.
SQ Sequence 566 AA;
Query Match 40.3%; Score 217; DB 1; Length 566;
Best Local Similarity 47.0%; Pred. No. 2.27e-09;
Matches 31; Conservative 15; Mismatches 16; Indels 4; Gaps 3;
DB 35 ERDPRQYECQRCRESATEEREOECQRCERY-KEQOROE--BELORQYQ-CQG 90

PI Bower NT, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI: 98-377279/32.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PR Useful for controlling microbial infestations of plants or mammals
PS Claim 1: Page 58-60: 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 593 AA;

Query Match 22.6%; Score 122; DB 1; Length 593;
Best Local Similarity 36.8%; Pred. No. 6.38e-02;
Matches 21; Conservative 12; Mismatches 20; Indels 4; Gaps 4;

DB 39 OCVRCEDEPRMHPORPCLEOC-REDEKREKERSREHADDRESGSSSEDEROEKEX 94
QY 132 HCQR-ETEPHMQT-CQQRCEKRYEKERKQKREDEQREDEKYE-ERKMEEDN 185

RESULT 10
ID W62836 standard; Protein: 33 AA.
AC W62836;
DE 27-OCT-1998 (first entry)
DE Zea mays antimicrobial protein.
KW antimicrobial protein; infestation; control.
OS Zea mays.
PN WO9827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997: AU0874.
PR 20-DEC-1996: AU-004275.
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bower NT, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI: 98-377279/32.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PR Useful for controlling microbial infestations of plants or mammals
PS Disclosure: Page 60: 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 33 AA;

Query Match 21.0%; Score 113; DB 1; Length 33;
Best Local Similarity 44.4%; Pred. No. 2.98e-01;
Matches 12; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

DB 6 ECRROCLRRHCGQPMETOCMRRCRR 32
QY 128 CQKHQR-ETEPHMQT-CQQRCEKRYEKERKQKREDEQREDEKYE-ERKMEEDN 185

RESULT 11
ID W90342 standard; Protein: 409 AA.
AC W90342;
DE 24-MAY-1999 (first entry)
DE G. max truncated SBP2 protein.
KW SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;
KM seed; carbohydrate content; soybean.
OS Glycine max.
PN WO9853086-A1.
PD 26-NOV-1998.
PF 21-MAY-1998: U10465.
PR 22-MAY-1997: US-047568.
PA (UNIM) UNIV WASHINGTON STATE RES FOUND.
PI Chao WS, Grimes HD;
DR WPI: 99-070155/06.
PT transgenic plant sucrose binding proteins - used to develop
PT transgenic plants which can have enhanced or decreased sucrose
PT uptake activity in developing seeds
PS Claim 7: Page 39-40: 58pp; English.
CC This sequence represents a novel sucrose binding protein, SBP2 isolated
CC from Glycine max. This protein is used in a method resulting in the
CC production of a modified plant sucrose binding protein (SBP) which has a
CC modified amino acid sequence compared to a corresponding wild-type SBP.

CC and where expression of the modified SBP in a yeast assay system confers
CC enhanced sucrose uptake compared to the corresponding wild-type SBP.
CC The products of the invention can be used for producing transgenic plants
CC which have modified sucrose uptake activity, particularly in developing
CC seeds. Enhanced sucrose uptake activity in developing seeds may be
CC desirable where it is an advantage to increase the carbohydrate content
CC of the seed (e.g. where the seed is the primary plant material harvested,
CC such as soybean). In contrast, decreased sucrose uptake activity in
CC seeds might be desirable where the vegetative material of the plant is
CC harvested. The SBP regulatory regions confer specific or enhanced
CC expression in developing seeds and so may be used to express any
CC transgene in developing seeds.
SQ Sequence 409 AA;

Query Match 21.0%; Score 113; DB 1; Length 409;
Best Local Similarity 34.5%; Pred. No. 2.98e-01;
Matches 20; Conservative 15; Mismatches 18; Indels 5; Gaps 3;

DB 42 CKHQCOQROQYTESKRTCLQCDSM--KQER-EKQVEETREKEEHQOHEEED 95
QY 129 CQKHQR-RETEPHMQT-CQQRCEKRYEKERKQKREDEQREDEKYE-ERKMEEDN 185

RESULT 12
ID W90341 standard; Protein: 489 AA.
AC W90341;
DE 24-MAY-1999 (first entry)
DE G. max SBP2 protein.
KW SBP1; sucrose binding protein; SBP2; sucrose uptake; transgenic plant;
KM seed; carbohydrate content; soybean.
OS Glycine max.
PN WO9853086-A1.
PD 26-NOV-1998.
PF 21-MAY-1998: U10465.
PR 22-MAY-1997: US-047568.
PA (UNIM) UNIV WASHINGTON STATE RES FOUND.
PI Chao WS, Grimes HD;
DR WPI: 99-070155/06.
PT transgenic plant sucrose binding proteins - used to develop
PT transgenic plants which can have enhanced or decreased sucrose
PT uptake activity in developing seeds
PS Claim 13b: Page 37-38: 58pp; English.
CC This sequence represents a novel sucrose binding protein, SBP2 isolated
CC from Glycine max. This protein is used in a method resulting in the
CC production of a modified plant sucrose binding protein (SBP) which has a
CC modified amino acid sequence compared to a corresponding wild-type SBP,
CC and where expression of the modified SBP in a yeast assay system confers
CC enhanced sucrose uptake compared to the corresponding wild-type SBP.
CC The products of the invention can be used for producing transgenic plants
CC which have modified sucrose uptake activity, particularly in developing
CC seeds. Enhanced sucrose uptake activity in developing seeds may be
CC desirable where it is an advantage to increase the carbohydrate content
CC of the seed (e.g. where the seed is the primary plant material harvested,
CC such as soybean). In contrast, decreased sucrose uptake activity in
CC seeds might be desirable where the vegetative material of the plant is
CC harvested. The SBP regulatory regions confer specific or enhanced
CC expression in developing seeds and so may be used to express any
CC transgene in developing seeds.
SQ Sequence 489 AA;

Query Match 21.0%; Score 113; DB 1; Length 489;
Best Local Similarity 34.5%; Pred. No. 2.98e-01;
Matches 20; Conservative 15; Mismatches 18; Indels 5; Gaps 3;

DB 42 CKHQCOQROQYTESKRTCLQCDSM--KQER-EKQVEETREKEEHQOHEEED 95
QY 129 CQKHQR-RETEPHMQT-CQQRCEKRYEKERKQKREDEQREDEKYE-ERKMEEDN 185

RESULT 13
ID R26941 standard; Protein: 316 AA.
AC R26941;
DE 08-FEB-1993 (first entry)

DE P.falciparum USA-R-NR protein.
KM Malaria; hepatocyte; sporozoite; plasmid DG 536; T-cell epitope;
KW paludism; liver stage-specific antigen.
OS Plasmodium falciparum.
FT key
FT location/Qualifiers
FT 1..209
FT /label=repeat_region
FT /note="contains 12 x 17mer repeats"
FT 210..316
FT /label=non-repeat_region
PN WO9213884-A.
PD 20-AUG-1992.
PF 05-FEB-1992; F00104.
PR 05-FEB-1991; FR-001286.
PI (INSP) INST PASTEUR.
PI Druilhe P, Guerin-Marchand C, Guerinmarchand C;
DR WPI: 92-299985/36.
DR N-PSDB: 028115.
PT Polypeptide(s) derived from liver stage of Plasmodium falciparum
PT - for vaccination against, treatment of and diagnosis of malaria
PS Disclosure; Fig 1; 81pp; French.
CC A genomic DNA bank of P.falciparum EcoRI fragments prepared in
CC lambda gtl1 was used to transform E.coli. The expression library was
CC screened with human antisera against antigens of all stages of P
CC falciparum. The library was rescreened with antibodies affinity-
CC purified on a clone which was able to recognise antibodies specific
CC to the hepatic phase. About 40 clones were detected which produced a
CC characteristic LSA epitope. The clone with the largest insert
CC (950 bases) encoded LSA-R-NR containing a 12-repeat region followed
CC by a non-repeat region. Preferred antigenic polypeptides of the
CC invention are derived from the amino acid sequence of LSA-R-NR.
SQ Sequence 316 AA;

Query Match 19.7%; Score 106; DB 1; Length 316;
Best Local Similarity 33.8%; Pred. No. 9,71e-01;
Matches 23; Conservative 13; Mismatches 30; Indels 2; Gaps 2;

DB 161 RAKEKLOE-QOSDLEORRAKEKLOEQSDLEQERR-AKKELOEQORDLEQRRADTKKRL 218
QY 118 RQDPQOQYEQCCKHCQRRETEPRHMOTCOQRCERYEKEKRRQKRYEQREDEKYE 177
DB 219 ERKKEHGD 226
QY 178 ERKKEEDN 185

RESULT 14
ID R26944 standard; Protein: 493 AA.

AC R26944;
DT 08-FEB-1993 (first entry)
DE P.falciparum LSA gene C-terminal region.
KM Malaria; hepatocyte; sporozoite; T-cell epitope; B-cell epitope;
KW paludism; liver stage-specific antigen.
OS Plasmodium falciparum.
FT key
FT location/Qualifiers
FT 13..213
FT /label=repeat_region
FT 214..493
FT /label=non-repeat_region
PN WO9213884-A.
PD 20-AUG-1992.
PF 05-FEB-1992; F00104.
PR 05-FEB-1991; FR-001286.
PI (INSP) INST PASTEUR.
PI Druilhe P, Guerin-Marchand C, Guerinmarchand C;
DR WPI: 92-299985/36.
DR N-PSDB: 028119.
PT Polypeptide(s) derived from liver stage of Plasmodium falciparum.
PT - for vaccination against, treatment of and diagnosis of malaria
PS Claim 2; Fig 8-10; 81pp; French.
CC The 3' part of the P.falciparum liver-stage specific antigen (LSA)
CC gene codes for a polypeptide sequence which carries a T cell epitope
CC characteristic of a protein produced in hepatocytes infected with

CC P.falciparum. The polypeptide can be used in the preparation of
CC vaccines against malaria.
SQ Sequence 493 AA;

Query Match 19.7%; Score 106; DB 1; Length 493;
Best Local Similarity 33.8%; Pred. No. 9,71e-01;
Matches 23; Conservative 13; Mismatches 30; Indels 2; Gaps 2;

DB 165 RAKEKLOE-QOSDLEORRAKEKLOEQSDLEQERR-AKEKLOEQORDLEQRRADTKKRL 222
QY 118 RQDPQOQYEQCCKHCQRRETEPRHMOTCOQRCERYEKEKRRQKRYEQREDEKYE 177
DB 223 ERKKEHGD 230
QY 178 ERKKEEDN 185

RESULT 15
ID W62841 standard; Protein: 28 AA.

AC W62841;
DT 27-OCT-1998 (first entry)
DE Stenocarpus sinuatus antimicrobial protein.
KW antimicrobial protein; infestation; control.
OS Stenocarpus sinuatus.
PN WO9827805-A1.
PD 02-JUL-1998.
PF 22-DEC-1997; AU0874.
PR 20-DEC-1996; AU-004275.
PI (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
PI Bower NJ, Goulter KC, Green JL, Manners JM, Marcus JP;
DR WPI: 98-377279/32.
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
PS Claim 1; Page 66; 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
SQ Sequence 28 AA;

Query Match 19.5%; Score 105; DB 1; Length 28;
Best Local Similarity 46.2%; Pred. No. 1.15e+00;
Matches 12; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

DB 2 DPIRQOQLCMRCQOQEKDPQOQOC 27
QY 121 DPQOQYEQCCKHCQRRETEPRHMOTC 146
Search completed: Sat May 13 09:43:35 2000
Job time : 10 secs.

